JOURNAL OF RESEARCH IN SCIENCE TEACHING VOL. 37, NO. 9, PP. 963–980 (2000)

The Effects of Professional Development on Science Teaching Practices and Classroom Culture

Jonathan A. Supovitz[1], Herbert M. Turner[2]

[1]*Consortium for Policy Research in Education, University of Pennsylvania, 3440 Market St. #560, Philadelphia, Pennsylvania 19104*

[2]*University of Pennsylvania*

Received 10 June 1999; accepted 30 May 2000

Abstract: The current science education reform movement emphasizes the importance of professional development as a means of improving student science achievement. Reformers have developed a vision for professional development based upon intensive and sustained training around concrete tasks that is focused on subject-matter knowledge, connected to specific standards for student performance, and embedded in a systemic context. Using data from a National Science Foundation Teacher Enhancement program called the Local Systemic Change initiative, this study employs hierarchical linear modeling to examine the relationship between professional development and the reformers' vision of teaching practice. The findings indicate that the quantity of professional development in which teachers participate is strongly linked with both inquiry-based teaching practice and investigative classroom culture. At the individual level, teachers' content preparation also has a powerful influence on teaching practice and classroom culture. At the school level, school socioeconomic status was found to influence practice more substantially than either principal supportiveness or available resources.© 2000 John Wiley & Sons, Inc. J Res Sci Teach 37: 963–980, 2000

Introduction

Staff development lies at the heart of nearly every educational effort to improve student achievement. Yet, paradoxically, the development of educators is a much maligned enterprise. In a 1985 national survey, teachers ranked in-service training as their *least effective* source of learning (Smylie, 1989). In his review of the literature on professional development, Guskey (1986) noted that nearly every major work on the topic of staff development disparaged its effectiveness. He attributed these historically dismal results to a poor understanding of teachers' motivations and a lack of insight into both the individual and environmental factors in the process of change. In an analysis of current professional development efforts, Little (1993)

Correspondence to: J. A. Supovitz. E-mail: jons@gse.upenn.edu
Contract grant sponsor: National Science Foundation
Contract grant number: RED-9255369

concluded that existing models of professional development are not adequate to achieve ambitious learning goals.

Although professional development may not have realized its potential, it is still seen as the best bet for changing teaching practices, because alternative methods, such as policies and programs that regulate teacher behavior, have fared no better (Smylie, 1996). Spillane and Thompson (1997) found that the capacity of local education policymakers was insufficient to support ambitious instructional reform. Other researchers have found that teachers often reshape or ignore policies that are intended to influence their basic classroom routines (Cohen & Ball, 1990; Wilson & Corbett, 1990). In a comprehensive review of research on alternative explanations for student achievement, Hawley and Rosenholtz (1984) concluded that ‘‘In virtually every instance in which researchers have examined the factors that account for student performance, teachers prove to have a greater impact than program. This is true for average students and exceptional students, for normal classrooms and special classrooms’’ (p. 3).

Despite historical shortcomings, evolutionary advances in science professional development hold promise as a way of influencing the teaching and learning of science in American public schools. Over the past decade, researchers and educators have forged a remarkable level of national consensus about what may constitute effective science professional development. They posit that professional development is most likely to be of high quality if it includes a set of six critical components. In this paper we use the term ‘‘high quality professional development’’ to refer specifically to teacher training experiences that incorporate most or all of these six elements.

First, high quality professional development must immerse participants in inquiry, questioning, and experimentation and therefore model inquiry forms of teaching (Arons, 1989; McDermott, 1990; Bybee, 1993). Little (1993) argues that these reforms ‘‘constitute a departure from canonical views of curriculum and from textbook-centered or recitation-style teaching’’ and represent ‘‘a substantial departure from teachers’ prior experience’’ (p. 130). Marek and Methaven (1991) found that the programs that model scientific reasoning have a greater influence on student achievement than did programs that taught teachers to use specific curricula.

Second, reformers argue that professional development must be both intensive and sustained (Smylie, Bilcer, Greenberg, & Harris, 1998; Hawley & Valli, 1999). The National Science Education Standards (NRC, 1996) call for more long-term, coherent professional development plans. Evidence on this point, though, is mixed. In a review of science and mathematics professional development programs that had evidence of impact on student achievement, Kennedy (1998) found little relationship between professional development contact time and student learning, although the association appeared to hold more in science than mathematics.

Third, staff development must engage teachers in concrete teaching tasks and be based on teachers’ experiences with students (Darling-Hammond & McLaughlin, 1995). Studies have shown that staff development undertaken in isolation from teachers’ ongoing classroom duties seldom have much impact on teaching practices or student achievement (Zigarmi, Betz, & Jennings, 1977). Lieberman (1995) argues that the definition of professional development must be expanded to include ‘‘authentic opportunities to learn from and with colleagues *inside* the school’’ (p. 591).

Fourth, professional development must focus on subject-matter knowledge and deepen teachers’ content skills (Cohen & Hill, 1998). The National Science Education Standards call for professional development to emphasize essential science content (NRC, 1996). Kennedy (1998) concluded that ‘‘programs that focus on subject matter knowledge and on student learning of particular subject matter are likely to have larger positive effects on student learning than are programs that focus on teaching behaviors’’ (p. 11).

Fifth, professional development must be grounded in a common set of professional development standards and show teachers how to connect their work to specific standards for student performance (NRC, 1996; Hawley & Valli, 1999). Work in cognitive development suggests that more complex knowledge and problem-solving skills require more sophisticated teaching strategies (Borko & Putnam, 1995), and that this kind of teaching can be achieved through setting higher learning goals (Resnick & Klopfer, 1989). One of the major findings from the Third International Mathematics and Science Study is that common, high standards are strongly related to national achievement (Schmidt, 1999).

Finally reform strategies must be connected to other aspects of school change (Fullan, 1991; O'Day & Smith, 1993; Corcoran & Goertz, 1995). One of the most persistent findings from school improvement research is the intimate relationship between staff development and school improvement. As Fullan (1991) states, "staff development cannot be separated from school development" (p. 331). According to Lieberman (1995), reform plans that have a chance to succeed must create a "culture of inquiry," which is an "ongoing part of teaching and school life." Marsh and LeFever (1997) found that school leadership is critical to school reform.

The Logic of Professional Development

The implicit logic of focusing on professional development as a means of improving student achievement is that high quality professional development will produce superior teaching in classrooms, which will, in turn, translate into higher levels of student achievement. School environments, as well as district and state policies, are powerful mediators of this sequence. This chain of logic is graphically represented in Figure 1.

Empirical evidence confirming this hypothesized chain of events in science is starting to emerge. A handful of recent studies have investigated the relationship between professional development and teaching practice. In a study of teachers who participated in Ohio's Statewide Systemic Initiative in science and mathematics, Supovitz, Mayer, and Kahle (2000) found that highly intensive (160 h), inquiry-based professional development changed teachers' attitudes towards reform, their preparation to use reform-based practices, and their use of inquiry-based teaching practices. Further, they found that these changes persisted several years after teachers concluded their experience. Allen and Lederman (1998) described a successful Teacher Academy for science and mathematics, which was based on intensive professional development

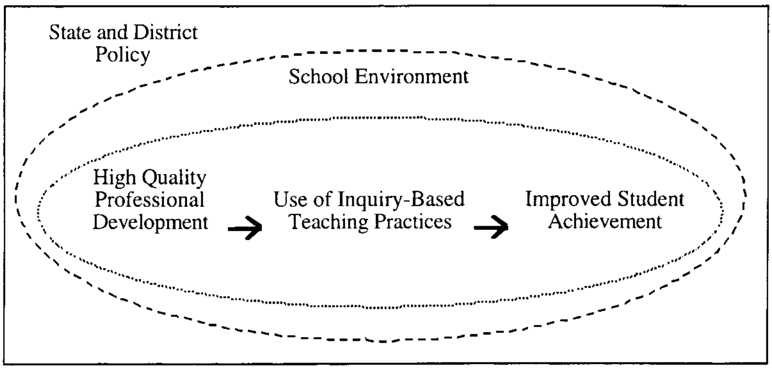

Figure 1. Model depicting theoretical relationship between professional development and student achievement.

(160–190 h over three years), high quality curriculum, and school leadership teams to provide school support and ensure institutionalization. While promising, both these cases showed the impact of high levels of professional development but were silent on the question of the appropriate level of professional development to change practice and how this could differ for teachers with different backgrounds.

Relatively few studies have documented a relationship between science teaching practices and student achievement in science. Kennedy (1998) sensibly noted this might be due to the "non-'3r' status" of science in the elementary school curriculum, which reduces the availability of science achievement data. An evaluation of the Merck Institute of Science Education by the Consortium for Policy Research in Education (CPRE, 1999) examined the link between inquiry-based teaching practices and student achievement on the open-ended portion of the Stanford 9 achievement test. The authors identified a statistical relationship between inquiry-based practice and fifth grade student achievement, but did not detect a similar association in the seventh grade. This was due, they argued, to more concerted implementation in the elementary school grades. Ohio's Statewide Systemic Initiative also has statistically linked their intensive professional development to gains in student achievement (Kahle & Rogg, 1996).

In this article, we contribute empirical evidence about the first of the two relationships depicted at the core of Figure 1—the relationship among high quality professional development and inquiry-based teaching practices and investigative classroom culture. We also investigate the influence of several aspects of the school environment on this relationship. In doing so we test several of the critical components hypothesized in reformers' construction of a definition of effective professional development. Using both teacher and principal survey data from a large-scale National Science Foundation (NSF) initiative called the Local Systemic Change (LSC) through Teacher Enhancement program, we explore the relationship between teacher background characteristics, teacher professional development experiences, school environment characteristics, and teacher practices and classroom culture.

More specifically, in this paper we address the following research questions:

1 Is high quality professional development that utilizes standards-based curriculum and is supported by a systemic context statistically related to inquiry-oriented teaching?
2. If so, what levels of professional development are associated with greater use of inquiry-based teaching practices?
3. How do teacher background characteristics mediate the relationship between professional development and practice?
4. What school demographic and environmental characteristics support or present barriers for inquiry-based teaching?

The Local Systemic Change Initiative

In 1995, the National Science Foundation's (NSF) Teacher Enhancement program funded the first cohort in a new initiative called the Local Systemic Change (LSC) initiative. Projects, typically funded for three to five years, focused on primary (K–8) and/or secondary (7–12) science and/or mathematics. To date, 72 projects are currently participating or have participated. This report analyzes data from the 24 projects that were participating in the K–8 science component of the initiative in 1997.

The LSC program enacts many facets of the vision of science reformers. The mission of the LSC program is to support improvements in the teaching of science, mathematics, and technology by implementing a teacher enhancement strategy founded upon proven, research-based models. Projects must have a shared comprehensive vision of science education reform incorporating national content, teaching, professional development, and assessment standards. The LSC initiative is also distinguished from previous teacher enhancement efforts by its emphasis on preparing teachers to implement designated exemplary instructional materials in the classroom. LSC projects are expected to align local policies and practices in support of their objectives. The goal of the program is to reach at least 80% of the teachers within each locality with a minimum of 100 h of professional development over the life of the project. Thus, the LSC program explicitly encourages professional development that is consistent with what we have earlier defined as high quality. For the purposes of this paper, we assume that the actual professional development that LSC sites deliver is consistent with this vision. Although there is most certainly variation in the implemented professional development, no attempts were made to model this variation for this study.

Participation in the LSC program also requires participation in the LSC core evaluation. NSF contracted with Horizon Research, Inc. (HRI) to design and oversee the core evaluation, which is carried out by local evaluators at each site. Each LSC evaluation, conducted annually, includes the surveying of a random sample of 300 teachers of science from the population of participating science teachers in each project and a survey of the population of participating school principals. Evaluation data also include observations of a small sample of classrooms and professional development activities; interviews with a sample of participants, principals, district administrators, and program staff; and additional data on the support system and environment of each project. For surveys, projects are required to achieve an 80% response rate. Each project may supplement its core data with additional, more customized, evaluation activities. Projects are not required to collect, nor report, student achievement data as part of the core evaluation. Unfortunately, this constrains the possibility of following the logic of professional development from teaching practices to student performance using these data.

Methods

This study is based upon the survey data collected from teachers and principals in 1997 as part of the core evaluation of the LSC initiative. The surveys asked teachers questions about their attitudes, beliefs, and teaching practices, as well as for personal demographic information. Principals of the schools in which these teachers worked were asked to answer questions about their support for the reform, as well as demographic data on the school and community.

Flora and Panter (1999) conducted both exploratory and confirmatory factor analyses on the 1996 and 1997 LSC teacher and principal survey data. They developed a series of distinct scales of inquiry-based teaching practice, investigative classroom culture, teachers' attitudes towards reform, teacher content preparedness, principal support, and classroom and school resource availability. Based on this work, individual scale values were constructed for each teacher or school. Missing values were imputed using multiple regression imputation. Cases were deleted from the sample where more than one-third of the items for a scale were missing.

Reliability and Validity of Self-Reported Teacher Survey Data

Educational researchers debate the reliability and validity of self-reported teacher survey data. On one hand, surveys are more cost-effective than observations, interviews, artifacts, or teacher logs as a way to collect data on teaching practices. On the other hand, it is unclear whether teachers can validly report on the schooling process and whether curricular practice can be accurately measured on a survey without observing the interactions between teachers and students (Burstein, McDonnell, Van Winkle, Ormseth, Mirocha, & Guitton, 1995).

Researchers in the 1990s learned much about the strengths and weaknesses of surveys to accurately collect data on teaching practices. Porter, Kirst, Osthoff, Smithson, and Schneider (1993) examined the consistency between survey responses pertaining to instructional style and detailed teacher logs describing actual lessons and concluded that ‘‘substantial’’ overlap existed. Burstein et al. (1995) used interviews, observations, daily teacher logs, and classroom artifacts to validate survey data. They found that survey data could validly depict topical content and instructional strategies, but that instructional goals were more difficult to accurately capture through survey research. Mayer (1999) explored the reliability and validity of survey data as part of his study of the relationship between NCTM-based teaching practices and student achievement. Assessing the reliability of his surveys, Mayer surveyed a group of teachers twice in a four-month period and found a correlation of .69 between responses on the two administrations. To validate his instruments, Mayer observed a random sample of nine classes and found a correlation of .85 between observational data and survey responses. Mullens (1998) described a National Center for Education Statistics pilot study that compared teacher survey responses to daily activity logs and classroom observations. He discovered that teachers tended to under-report on questionnaires on both the frequency with which they employed certain learning objectives and classroom instructional activities.

The Sample

Our data, survey responses from teachers and principals in 24 communities from across the United States, represent a rare national view of science teaching and support. The localities ranged from large urban areas like San Francisco and Chicago, to smaller cities like Waltham, Massachusetts and Mesa, Arizona, to small towns like Elizabeth City, North Carolina and Midland, Michigan. Depending on the size of the project, the potential population of science teachers ranged from 2027 in the largest project to 173 in the smallest one.[1] Regardless of the size of the community, as part of the core evaluation, HRI randomly sampled 300 teachers of science from the population of science teachers in each LSC project.

In 787 schools, 4903 teachers completed LSC surveys in the spring of 1997. After removing invalid responses (mostly due to either missing data that could not be reasonably imputed or missing linking information between teacher and principal surveys), we were left with surveys from 3464 science teachers and 666 principals in the 24 localities. These formed the basis for our analyses. To account for differences in the probabilities that teachers in projects of different sizes were in the sample, weights were allocated to each respondent.

Another unique quality of these data were teachers’ range of exposure to professional development. Since projects had different strategies for reaching the required 80% of their science teachers—for example, some provided training to teachers in a set of grades each year, some served a proportion of teachers each year, and some relied on volunteers and provided incentives—and were at varying stages of completion, different proportions of their science teachers had participated in professional development at the time these data were collected. The random sample thus resulted in a representative sample of teachers from the participating projects that had participated in professional development to varying degrees. While some teachers had participated intensely in professional development, others had only slight involvement, and still others had no LSC professional development at all. Thus these data are a rare representative sample of teachers with varying levels of formal exposure to high quality professional development.

Outcome Variables

Surveys of teachers asked in detail about their attitudes about teaching, their classroom practices, and their LSC professional development experiences (if any to date). In their factor analysis of the LSC survey data, Flora and Panter (1999) discovered that the survey items that asked teachers about the range of their classroom pedagogical practices cleaved nicely into several distinct factor groupings. One group of items depicted investigative classroom practices, while another focused more on the culture of the classroom. Two practice-related scales (i.e., the "reform indicators") were used as outcome measures for this study.

The composite of teachers' investigative practices was based on a series of questions about the frequency of their use of a number of reform-based teaching practices, including having students "*engage in hands-on activities; design or implement their own investigation; write reflections in a notebook or journal; and work on extended science investigations or projects*." Teacher responses were on a five-point frequency scale (from Never to Almost Daily). The reliability of the investigative practice scale, computed by using Cronbach's internal consistency measure coefficient alpha, was .82.

The composite of teachers' classroom culture of investigation was based on questions about their classroom strategies when teaching science. For example, teachers were asked how frequently they did such things as "*arrange seating to facilitate student discussion; require students to supply evidence to support their claims; encourage students to explain concepts to one another; and have students work in cooperative groups*." Teachers again were asked to respond on a five-point frequency scale. The reliability of this scale was .88.

A complete list of the questions that comprised the scales and their reliabilities are shown in the Appendix. All scales were standardized with a mean of 0 and a standard deviation of 1. This allowed for both easier interpretation as well as a way of interpreting the scales relative to each other.

Predictor Variables

Predictor variables used in this study were based upon information collected from both teachers and school principals. Table 1 shows the demographic characteristics of the teachers in the sample. The teachers typically were white, female, elementary school teachers with around eight years of teaching experience. Forty-four percent of the sample had not participated in LSC professional development. About 40% had participated in less than 40 h of LSC professional development, and the remaining 20% had participated in considerably more than 40 h of LSC professional development.

Four additional individual teacher variables were included in the model: teachers' attitudes towards reform, their content preparation, their perceptions of principal support, and their available resources. A scale of teachers' attitudes towards reform was included in the model to adjust for differences in the reform indicators that were due to different levels of teacher advocacy of the reforms. The scale used to develop a teacher's attitude towards inquiry-oriented teaching contained 10 Likert scale questions. These asked teachers about the importance of such activities as *engaging students in inquiry-oriented activities, having students work in cooperative learning groups, and using informal questioning to assess student understanding*. The reliability of this scale was .81.

To assess the influence of administrative support on teaching practices and classroom culture, a scale was constructed from nine items that asked teachers about the support of their principal. Items in this scale included the extent of teacher agreement to such statements as *My*

Table 1
Individual characteristics of potential LSC participants

	Science teachers (n = 3464)
Teacher gender	
Female	88%
Male	12%
Teacher ethnicity	
White	77%
Minority	23%
Grade levels	
K–6 Teachers	94%
7–8 Teachers	06%
Years of teaching experience	(5 point scale: 0–2, 3–5, 6–10, 11–20, 21+)
Mean	3.43
Standard deviation	1.35
Minimum	1
Maximum	5
Quantities of professional development	
None	44%
1–19 h	26%
20–39 h	12%
40–79 h	10%
80–159 h	06%
More than 160 h	04%

principal encourages the implementation of current national standards in science education, My principal encourages innovative instructional practices, and *My principal acts as a buffer between teachers and external pressures*. The reliability of this scale was .89.

The content preparation scale measured how well-prepared teachers felt to teach 11 science topics that are commonly taught in elementary school. These included mixtures and solutions, electricity, sound, anatomy, and ecology. The reliability of the content preparation scale was .91. The final individual level scale was teacher resource availability. This scale included questions on the time available for teachers to prepare lessons, opportunities for teachers to work with other teachers, and opportunities for teacher professional development. The reliability of this four-item scale was .85.

Several attributes which characterized the 666 schools in our sample were included in the model as well. These included the size of the school, the type of community in which the school resided, and the proportion of students in each school who received free or reduced lunch. Table 2 shows the descriptive statistics for these school-level variables. School sizes ranged widely; the smallest having just 121 students, while the largest school in the sample contained about 1500 students. On average, schools had just over 500 students. Schools also varied widely in the proportions of their students receiving free or reduced lunch, which is considered to be a proxy for the socioeconomic status of the school. In the average school, about half of the students receive lunch assistance, with a standard deviation of 31%.

Table 2
School characteristics of potential LSC participants

	Schools (n = 666)
School size	
Mean	535
Standard deviation	227
Minimum	121
Maximum	1449
Students in school on lunch assistance	
Mean	48%
Standard deviation	31%
School community locations	
Rural area	08%
Town	17%
Suburban area	24%
Urban area	51%

The final school-level variable included in the models was a scale developed from principal survey responses about the resources available to the school. The school resource scale included such things as available instructional materials, funds for purchasing science equipment and supplies, the system of managing instructional resources at the district or school level, and the importance the school places on science.

Although not a nationally representative sample, the LSC sample was in many ways similar to teachers nationwide. Table 3 shows the demographics of the LSC sample in comparison to national estimates of elementary school teachers reported by the National Center for Educational Statistics (1995). The LSC sample consisted of teachers of science, while the national estimates represented all elementary (K–8) teachers. The LSC sample had 16% greater female and 10% greater minority than the national sample. Teachers in the LSC sample were slightly more experienced than were teachers in the national sample. The school sizes of the two samples were similar, but the LSC projects were more likely to be located in urban areas in comparison to the community distribution of schools nationally. Thus the LSC sample was similar in terms of school size and teaching experience, but was more urban, female, and had a greater proportion of minority teachers than was the distribution of teachers across the nation.

Analytic Strategy

To investigate the relationship between professional development and the reform indicators of inquiry-based teaching practices and investigative classroom culture, we used a series of hierarchical linear models (HLM). Since teachers are naturally grouped within schools and we were interested in both the effects of individuals' characteristics as well as their school environments, HLM was an appropriate statistical technique to more precisely model these nested relationships (Bryk & Raudenbush, 1992). The models were estimated using the HLM/3L software of Bryk, Raudenbush, and Congdon (1996). The HLM/3L software can include predictor variables associated with individual teachers and schools, and incorporate variation occurring distinctly at each level that is discrete from measurement error variance.

Table 3
Comparison of LSC sample teacher and school characteristics and National estimates of elementary teacher and school characteristics

	National Estimates 1994	LSC Sample 1997
Teacher characteristics		
Gender		
Female	72%	88%
Male	28%	12%
Ethnicity		
White	87%	77%
Minority	13%	23%
Median years of teaching		
Experience	15	11
School characteristics		
Average school size	515	535
Community location		
Rural	29%	8%
Town	27%	17%
Suburb	21%	24%
Urban	23%	51%

The analysis was done with two-level hierarchical models. The models linked the reform indicators to teacher and school characteristics using two levels of statistical models. The "level-1" model expresses one of the reform indicators as a function of individual teacher demographics and perceptions of school characteristics. The "level-2" model expresses the parameters from the level-1 model as a function of the school characteristics within which that teacher resides. To facilitate interpretation, all variables were entered grand mean centered, so that model coefficients represent the mean values for the entire sample. While we tested several interaction terms both between level-1 characteristics and between level-1 and level-2 characteristics, there were no interactions worth including in the final models.

Results

Our main question of interest in this study is the relationship between professional development and the reform indicators. Tables 4 and 5 show the results of the HLM models of the relationship between teacher and school characteristics, professional development, and inquiry-based teaching practices and investigative classroom culture. In both models, increasing amounts of professional development were statistically associated with both greater teacher use of inquiry-based teaching practices and higher levels of investigative classroom culture.

In the model predicting inquiry-based teaching practices (Table 4), after adjusting for differences between teachers and schools, the results indicate that, on average, teachers with no professional development were predicted to employ inquiry-based practices four-tenths of a standard deviation less frequently than that of the average teacher in the sample. In fact, teachers with less than 40 h of professional development had more traditional practices (i.e., less inquiry-oriented) than did the average teacher. Teachers with between 40 and 79 h of professional

Table 4
Coefficients for individual and school characteristics in HLM model predicting inquiry-based teaching practices

	Coefficient	Standard Error
Level 1 Variables		
Intercept	.021	.021
No professional development	−.409***	.049
1–19 h of professional development	−.193***	.050
20–39 h of professional development	−.077	.055
80–159 h of professional development	.160*	.076
More than 160 h of professional development	.168~	.103
Middle School (7–8) teacher	−.141~	.087
Male teacher	−.062	.049
Minority teacher	.197***	.033
Years of teaching experience	.002	.010
Teacher's attitude towards reform	.138***	.014
Content preparation	.220***	.014
Principal's supportiveness	.083***	.015
Classroom resource availability	.038**	.015
Level 2 Variables		
Proportion of students in school on free or reduced Lunch	−.222**	.081
Rural community	.128	.085
Town	−.137*	.062
Suburban community	−.068	.059
School resource availability	−.030	.020
School size (per 100 students)	−.018*	.009

***$P < 0.001$ **$P < 0.01$ *$P < 0.05$ ~$P < 0.10$

development (the omitted, or referent, group and thus the value of the intercept) had about average teaching practices. It was only after approximately 80 h of professional development that teachers reported using inquiry-based teaching practices significantly more frequently—about two-tenths of a standard deviation—than the average teacher.

The model detailed in Table 5 shows a similar relationship between professional development and classroom investigative culture. Each additional block of professional development was associated with a more investigative classroom culture, after adjusting for differences both between teachers and their schools. Similar to the investigative practices model, teachers with minimal exposure to professional development had below average inquiry-oriented classroom cultures, and it was only after 40–79 h that teachers' investigative classroom culture was above that of the average teacher.

The pattern of the relationship between different levels of professional development and the reform indicators is graphically shown in the two panels in Figure 2. The points in each panel show the predicted use of inquiry-based teaching practices and classroom investigative culture for values for an "average" teacher from an "average" school with different levels of professional development.[2] The curves indicate that increasing amounts of LSC professional development were strongly linked with increasing teacher use of inquiry-based practice and investigative classroom culture.

Table 5
Coefficients for individual and school characteristics in HLM model predicting investigative classroom culture

	Coefficient	Standard Error
Level 1 Variables		
Intercept	.016	.019
No professional development	−.326***	.049
1–19 h of professional development	−.209***	.050
20–39 h of professional development	−.129	.054
80–159 h of professional development	.022	.076
More than 160 h of professional development	.115~	.060
Middle School (7–8) teacher	−.035	.086
Male teacher	−.171***	.050
Minority teacher	.060~	.033
Years of teaching experience	−.031**	.010
Teacher's attitude towards reform	.204***	.014
Content preparation	.200***	.014
Principal's supportiveness	.063***	.015
Teacher resource availability	.020	.014
Level 2 Variables		
Proportion of students in school on free or reduced Lunch	−.324***	.076
Rural community	−.171*	.079
Town	.024	.058
Suburban community	.057	.055
School resource availability	.006	.019
School size (per 100 students)	.005	.008

*** $P < 0.001$ ** $P < 0.01$ * $P < 0.05$ ~ $P < 0.10$

Examination of the individual teacher characteristics revealed several differences in both practices and classroom culture of teachers associated with different background characteristics. Interestingly, male teachers were more traditional in both their investigative practices and in the culture of their classroom than female teachers. However, these gender differences were statistically significant (at the .05 level) only in the model of investigative classroom culture. Minority teachers, who represented 23% of the sample, reported significantly more frequent use of inquiry-based classroom practices and investigative classroom culture than did their White counterparts. Teaching experience was negatively associated with investigative culture, but was not related to inquiry-based teaching practices.

The most powerful individual influences on both teaching practices and investigative culture were teachers' content preparation and attitudes towards reform. Each standard deviation of increased content preparation a teacher reported was associated with a 20% increase in the use of both investigative teaching practices and investigative classroom culture. Further, the lack of an interaction between content preparation and professional development indicated that the magnitude of the effect of content preparation on practice was the same regardless of the intensity of teachers' professional development experiences.

Finally, and not surprisingly, teachers with more sympathetic attitudes towards reform used inquiry-based practices significantly more frequently and had more investigative classroom cultures than did more skeptical teachers. Teachers' attitudes towards reform were included in the models to control for this powerful effect.

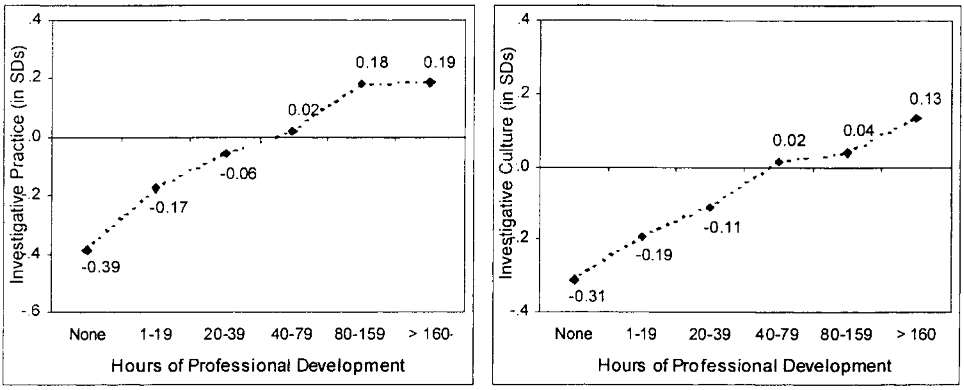

Figure 2. The relationship between different levels of professional development and classroom practice and culture.

Several school factors also played important roles in the extent that teachers used inquiry-based practices and developed investigative classroom cultures. One was the supportiveness of the school's principal. On average, teachers who felt supported by their principal reported significantly greater use of reform approaches than did teachers who did not feel encouraged by their school leader. Schools' available resources, as reported by the principal, which included instructional materials, time for teachers to plan and prepare lessons, and availability of science-relevant supplies, also had a statistically significant influence on teachers' investigative practices, but not on their investigative classroom culture.

The largest school-level influence on teachers' practices and classroom culture was school poverty, here defined as the percentage of students in a school on lunch assistance. Teachers in schools with high proportions of students on free and reduced lunch had, on average, significantly lower levels of both investigative culture and inquiry-based practices. For each additional standard deviation of students receiving lunch assistance, all else being equal, teachers used inquiry-based practices 20% less frequently and had about 30% less frequent use of investigative culture.

Finally, the type of community in which schools resided had relatively little influence on teachers' reform practices; teachers performed similarly in towns as well as urban, rural, and suburban areas. The only exception to this general pattern was that teachers in rural communities had significantly lower—by almost a quarter of a standard deviation, on average—levels of classroom investigative culture. In weighing the meaning of this result, however, it should be noted that teachers in rural areas represented only 8% of the sample.

Discussion

At first glance, the central finding of this study contradicts the historical skepticism about the efficacy of professional development. This work demonstrates a strong and significant relationship between professional development and teachers' practices and classroom cultures. But upon deeper reflection, these results are consistent with past findings. Most of the professional development experiences examined in previous studies were shorter in length and intensity, and their relative ineffectiveness is remarkably consistent with the effects of shorter professional development experiences in this sample. Dramatic results emerged when the experiences were deeper and more sustained. Both teaching practices and classroom cultures were affected most deeply after intensive and sustained staff development activities. In these analyses, it was only teachers with more than two weeks of professional development who reported teaching practices and classroom cultures above average. Further, it appears that it was somewhat more difficult to change classroom culture than teaching practices; the big change in teaching practice came after 80 h of professional development, while the big change in investigative culture came only after 160 h.

Standards and curriculum may have been the silent partners of professional development in influencing teaching practices and classroom cultures. All the LSC projects have a heavy standards emphasis and are required to use NSF-approved curriculum materials in support of their initiatives. By providing teachers with a more coherent series of grade-level appropriate, content-rich activities—often framed with helpful procedures and linked to larger science concepts as well as sequenced to meet national standards—even teachers with no professional development have access to these curriculum materials that encourage inquiry and investigative cultures. This might have affected their practices, raising the average, and resulting in an underestimation of the effects of professional development. Further research should try to explore the relative contributions of standards, curriculum, and professional development on teaching practices.

Content preparation was by far the most powerful individual teacher factor included in the models. Regardless of the extent of their professional development, there was a four-tenths of a standard deviation difference in practice between teachers with one standard deviation below average content preparation and teachers with one standard deviation above average content preparation. This finding is in agreement with earlier empirical works (Kennedy, 1998; CPRE, 1999; Cohen & Hill, 1998) and reinforces the emerging consensus about the critical importance of content knowledge in science teaching.

Several school factors were also important predictors of reform teaching. All other things being equal, teachers from poorer schools (i.e., those with higher proportions of students on free and reduced lunch) appeared to use more traditional teaching practices than did their counterparts from schools with more wealthy student populations. This finding might be associated with differences in class size, discipline, time allocations, or other characteristics that differentiate low and high poverty schools, but were not accounted for in the models. This bears further investigation. Also, and surprisingly, neither school resources nor principal supportiveness had powerful influences on teachers' uses of inquiry-based practices and investigative classroom culture.

Thus, the results of this investigation point towards a strong relationship between high quality professional development and the kinds of teaching practices that are advocated by science reformers. The fact that the teachers in the LSC sample are individually similar to teachers nationally, but come from more urban schools, suggests that these results can be replicated nationally because urban sites are generally considered more resistant to reform.

However, these analyses fail to capture several dimensions of real life that would confirm that an investment in intensive and sustained professional development, focused on subject-matter knowledge, connected to student work, and embedded in a systemic context and supportive school culture, is a sound one.[3] First, these models test only the first part of the hypothesis raised in Figure 1—the link between professional development and teaching practice. A more comprehensive test of the theory would also examine the relationship between teaching practice and student achievement. Do these types of teaching practices lead to higher levels of student achievement? Second, the data underlying these models are cross-sectional rather than longitudinal, and thus they cannot speak to the crucial question of whether professional development is linked to *changes* in practice. Further, the ability to model change in practice would allow us to look at whether professional development was increasing teachers content preparation as well. Does high quality professional development *change* practice? Third, these analyses assume that the quality of all of the professional development provided by the LSC's is equal and high. In fact, variation in the quality of the professional development within the six dimensions laid out in the introduction to this paper is likely. How does the variation in implementation affect practice? And finally, these analyses do not capture variation in the programmatic strategies of the LSC's. For example, some programs may emphasize coaching models, while others may make extensive use of study groups, and still others may utilize electronic support networks. Do different support strategies produce differential results? These questions suggest fertile ground for subsequent analyses.

Appendix: Survey Questions and Scales Used to Create Science Composites

Outcome Variable Scales

Teachers' Use of Investigative Teaching Practices (Reliability = .82). All questions on a 5 point scale: Never, Once or twice a semester, once or twice a month, once or twice a week, almost daily. How frequently do students: (a) make formal presentations to class, (b) engage in hands-on science activities, (c) design or implement their own investigations, (e) work on models or simulations, (f) work on extended science investigations or projects (a week or more in duration), (g) Participate in field work, (h) write reflections in a notebook or journal, (i) work on a portfolio.

Teachers' Use of Investigative Classroom Culture (Reliability = .88). All questions on a 5 point scale: Never, Once or twice a semester, once or twice a month, once or twice a week, almost daily. (a) Arrange seating to facilitate student discussion, (b) use open-ended questions, (c) require students to supply evidence to support their claims, (d) encourage students to explain concepts to one another, (e) encourage students to consider alternative explanations, (f) participate in discussions with the teacher to further science understanding, (g) work in cooperative learning groups, (h) share ideas or solve problems with each other in small groups.

Predictor Variable Scales

Teachers' Attitudes Towards Reform (Reliability = .81). All questions on a 5 point scale ranging from Strongly Disagree to Strongly Agree. (a) Provide concrete experience before abstract concepts, (b) develop students' conceptual understanding of science, (c) make connections between science and other disciplines, (d) have students work in cooperative learning groups, (e) have students participate in appropriate hands-on activities, (f) engage students in inquiry-oriented activities, (g) Use computers, (h) engage students in applications of science in a variety of contexts, (i) use portfolios, (j) use informal questioning to assess student understanding.

Teachers' Science Content Preparedness (Reliability = .91). All questions on a 5 point scale ranging from Not Well Prepared to Very Well Prepared. How well prepared do you feel to do each of the following? (a) The human body, (b) ecology, (c) rocks and soil, (d) astronomy, (e) processes of change over time (e.g., evolution), (f) mixtures and solutions, (g) electricity,

(h) sound, (i) forces and motion, (j) machines, (k) engineering and design principles (e.g., structures, models).

Teachers' Perceptions of Principal Support (Reliability = .89). All questions on a 5 point scale ranging from Strongly Disagree to Strongly Agree. (a) My principal encourages me to select science content and instructional strategies that address individual students' learning, (b) my principal accepts the noise that comes with an active classroom, (c) my principal encourages the implementation of current national standards in science education, (d) my principal encourages innovative instructional practices, (e) my principal enhances the science program by providing me with needed materials and equipment, (f) my principal provides time for teachers to meet and share ideas with one another, (g) my principal encourages me to observe exemplary science teachers, (h) my principal encourages teachers to make connections across disciplines, (i) my principal acts as a buffer between teachers and external pressures.

Teacher Resource Availability (Reliability = .85). All questions on a 5 point scale ranging from inhibits effective instruction to encourages effective instruction. (a) Time available for teachers to plan and prepare lessons, (b) opportunities for teachers to work with other teachers, (c) opportunities for teacher professional development, (d) importance the school places on science.

School Resource Availability (Reliability = .85). All questions on a 5 point scale ranging from inhibits effective instruction to encourages effective instruction. (a) Quality of available instructional materials, (b) access to computers for science instruction, (c) funds for purchasing equipment and supplies for science, (d) system of managing instructional resources at the district or school level, (e) time available for teachers to plan and prepare lessons, (f) opportunities for teachers to work with other teachers, (g) opportunities for teacher professional development, (h) importance that the school places on science, (i) consistency of science reform efforts with other school/district reform efforts, (j) public attitudes toward reform.

We would like to thank Iris Weiss and the staff at Horizon Research, Inc. for their assistance compiling the data for this research and their thoughtful critique of our analyses. All opinions, findings, and conclusions expressed in this article are those of the authors.

Notes

[1]In this paper we use the terms "teachers of science" and "science teachers" interchangeably. When we do so, we are referring to teachers who either teach science as part of their responsibilities or as their sole responsibility. These may include general elementary teachers who teach science in addition to other subjects or intermediate and middle school teachers who are science specialists.

[2]The values for this "average" teacher are derived simply by substituting the sample average values for each variable.

[3]It should be noted that these are weaknesses in these analyses, not weaknesses in the theory of the reformers.

References

Allen, E.E., & Lederman, L. (1998). Lessons learned: The teachers academy for mathematics and science. Phi Delta Kappan, October, 158–164.

Arons, A.B. (1989). What science should we teach? In The Biological Science Curriculum Study (Ed.), A BSCS Thirtieth Anniversary Symposium: Curriculum Development for the Year 2000 (pp. 13–20). Colorado Springs, CO: BSCS.

Borko, H., & Putnam, R.T. (1995). Expanding a teacher's knowledge base: A cognitive psychological perspective on professional development. In T.R. Guskey & M. Huberman (Eds.), Professional development in education: New paradigms & practices (pp. 35–66). New York: Teachers College, Columbia University.

Bryk, A.S., & Raudenbush, S.W. (1992). Hierarchical linear models: applications and data analysis methods. Newbury Park: Sage Publications.

Bryk, A.S., Raudenbush, S.W., & Congdon, R.T. (1996). Hierarchical linear and nonlinear modeling with HLM/2L and HLM/3L programs. Chicago: Scientific Software Inernational.

Burstein, L., McDonnell, L.M., Van Winkle, J., Ormseth, T., Mirocha, J., & Guitton, G. (1995). Validating national curriculum indicators. Santa Monica: Rand Corporation.

Bybee, R.W. (1993). Reforming science education: Social perspectives and personal reflections. New York: Teacher's College Press.

Cohen, D.K., & Ball, D.L. (1990). Policy and practice: An overview. Educational Evaluation and Policy Analysis, 12(3), 233–239.

Cohen, D.K., & Hill, H.C. (1998). State policy and classroom performance: Mathematics reform in California. CPRE Policy Brief. Consortium for Policy Research in Education.

Consortium for Policy Research in Education. (1999). A Close look at effects on classroom practice and student performance: A report of the fifth year of the Merck Institute for Science Education 1997–98. CPRE evaluation report: Author.

Corcoran, T., & Goertz, M. (1995). Instructional capacity and high performance. Educational Researcher, 24(9) 27–31.

Darling-Hammond, L., & McLaughlin, M.W. (1995). Policies that support professional development in an era of reform. Phi Delta Kappan, April 1995.

Flora, D.B., & Panter, A.T. (1999). Technical report: Analysis of the psychometric structure of the LSC surveys. Chapel Hill, NC: L.L. Thurstone Psychometric Lab.

Fullan, M. (1991). The new meaning of educational change. New York: Teachers College Press.

Guskey, T.R. (1986). Staff development and the process of teacher change. Educational Researcher, May, 1986.

Hawley, W.D., & Rosenholtz, S. (1984). Good schools: A synthesis of research on how schools influence student achievement. Peabody Journal of Education, 4, 1–178.

Hawley, W.D., & Valli, L. (1999). The essentials of effective professional development: A new consensus. In G. Sykes & L. Darling-Hammond (Eds.), Handbook of teaching and policy. New York: Teachers College.

Kahle, J.B., & Rogg, S.R. (1996). A pocket panorama of the landscape study, 1995. Miami University, Oxford, OH.

Kennedy, M.M. (1998). The relevance of content in inservice teacher education. Paper presented at the annual meeting of the American Educational Research Association. April, 1998. San Diego, CA.

Lieberman, A. (1995). Practices that support teacher development. Phi Delta Kappan, April, 1995.

Little, J.W. (1993). Teachers' professional development in a climate of educational reform. Educational evaluation and policy analysis. 15(2), 129–151.

Marek, E.A., & Methaven, S.B. (1991). Effect of the learning cycle upon student and classroom teacher performance. Journal of Research in Science Teaching, 28(1), 41–53.

Marsh, D.D., & LeFever, K. (1997). Educational leadership in a policy context. What happens when student performance standards are clear? Paper presented at the annual meeting of the American Educational Research Association. Chicago.

Mayer, D. (1999). Measuring instructional practice: Can policymakers trust survey data? Educational Evaluation and Policy Analysis, 21(2), 29–45.

McDermott, L.C. (1990). A perspective in teacher preparation in physics and other sciences: the need for special science courses for teachers. American Journal of Physics, 58.

Mullens, J.E. (1998). Validating teacher responses on instructional practice self-report surveys: Findings from a recent fieldtest. Paper presented at the annual conference of the Association for Public Policy Analysis and Management. New York: October 31, 1998.

National Center for Educational Statistics. (1995). Digest of Education Statistics 1995. Washington, DC: US Department of Education.

National Research Council. (1996). National Science Education Standards. Washington, DC: National Academy Press.

O'Day, J., & Smith, M. (1993). Systemic reform and educational opportunity. In S.H. Fuhrman (Ed.), Designing coherent education policy: Improving the system (pp. 250–312). San Francisco: Jossey-Bass.

Porter, A.C., Kirst, M.W., Osthoff, E.J., Smithson, J.L., & Schneider, S.A. (1993). Reform up close: An analysis of high school mathematics and science classrooms. Madison: Wisconsin Center for Educational Research.

Resnick, L., & Klopfer, L. (1989). Toward the thinking curriculum: An overview. Toward the thinking curriculum: Current cognitive research, 1989 yearbook of the Association for Supervision and Curriculum Development (pp. 1–10). Alexandria, VA: ASCD.

Schmidt, W. (1999). Presentation at the annual conference of the National Center for Education and The Economy, January, 8. San Diego, CA.

Smylie, M.A. (1989). Teachers' views of the effectiveness of sources of learning to teach. The Elementary School Journal, 89(5), 543–558.

Smylie, M.A. (1996). From bureaucratic control to building human capital: The importance of teacher learning in education reform. Educational Researcher, 25(9), 9–11.

Smylie, M.A., Bilcer, D.K., Greenberg, R.C., & Harris, R.L. (1998). Urban teacher professional development: A portrait of practice from Chicago. Paper presented at the annual meeting of the American Educational Research Association. April, 1998. San Diego, CA.

Spillane, J.P., & Thompson, C.L. (1997). Reconstructing conceptions of local capacity: The local education agency's capacity for ambitious instructional reform. Educational Evaluation and Policy Analysis, 19(2), 185–203.

Supovitz, J.A., Mayer, D.P., & Kahle, J.B. (2000). Promoting inquiry-based instructional practice: The longitudinal impact of professional development in the context of systemic reform. Educational Policy, 14(3), 331–356.

Wilson, B.L., & Corbett, H.D. (1990). Statewide testing and local improvement: An oxymoron? In J. Murphy (Ed.), The educational reform movement of the 1980s (pp.243–263). Berkeley, CA: McCutchan.

Zigarmi, P., Betz, L., & Jennings, D. (1977). Teachers' preferences in and perceptions of inservice. Educational Leadership, 34, 545–551.